Discovering relevancies in very difficult regression problems: applications to sensory data analysis

Jorge Díez and **Gustavo F. Bayón** and **José R. Quevedo** and **Juan José del Coz** and **Oscar Luaces** and **Jaime Alonso** and **Antonio Bahamonde**[1]

Abstract. Learning preferences is a useful tool in application fields like information retrieval, or system configuration. In this paper we show a new application of this Machine Learning tool, the analysis of sensory data provided by consumer panels. These data sets collect the ratings given by a set of consumers to the quality or the acceptability of market products that are principally appreciated through sensory impressions. The aim is to improve the production processes of food industries. We show how these data sets can not be processed in a useful way by regression methods, since these methods can not deal with some subtleties implicit in the available knowledge. Using a collection of real world data sets, we illustrate the benefits of our approach, showing that it is possible to obtain useful models to explain the behavior of consumers where regression methods only predict a constant reaction in all consumers, what is unacceptable.

1 ANALYSIS OF SENSORY DATA

An important part of the success of food industries relies on their ability to produce their specialties satisfying the consumers' sensory requirements. A survey of the use of sensory data in the food industry can be found in [8]; for a Machine Learning perspective, see [4].

Sensory data include the assessment of food products provided by two different kinds of groups of people usually called *panels*. The first one is made up of a small group of expert, trained judges; these will describe each product by attribute-value pairs. Here we must assume that a rating of "7" (in say, texture) means the same for a given expert in every product, though not necessarily for every expert. The second kind of panel is made up of untrained consumers; these are asked to rate their degree of acceptance or satisfaction about the tested products on a scale. The aim is to be able to relate sensory descriptions (human and mechanical) with consumer preferences in order to improve production decisions.

If we consider the whole data collected in a sensory study, we have to take into account that these data sets have some important properties that must be considered. First, we observe that the assessments come from a set of different consumers. This implies that we will have different scales in the available ratings. In other words, "7" does not mean the same for everybody. Second, in each *testing session* a small set of products is shown to consumers, but their ratings can not be considered in an absolute way; in fact, they use the assessments to express a relative ordering of the samples presented in that testing session. This is known as *batch effect*: an object presented in a batch with clearly worse objects will probably obtain a higher rating than if it were surrounded by preferable objects. Finally, consumers do not test all available samples. Typically, each consumer only participates in one or a small number of testing sessions.

Traditionally the process given to these data sets includes testing some statistical hypothesis [9, 8]. In all cases these previous approaches demand that all available food products must be rated by all consumers. An alternative approach can be based on regression.

In the next section we will present our approach to deal with sessions explicitly. The overall idea is avoid trying to predict the exact value of consumer ratings; instead we will look for a function that returns higher values to those products with higher ratings. Such functions are called preference or ranking functions [2].

2 BINARY SEPARATION AND PREFERENCES

Although there are other approaches to learn preferences, following [5, 7, 1] we will try to induce a real *preference* or *ranking function* that maximizes the probability of assigning a higher rating to an object v than to an object u whenever v is preferable to u. We have a collection of preference judgments $PJ = \{v_j > u_j : j = 1, \ldots, m\}$, and following [5] we are looking for a function such that

$$F(v_j, u_j) > 0 \text{ and } F(u_j, v_j) < 0 \quad \forall j = 1, \ldots, m. \tag{1}$$

Therefore, we have a binary classification problem that can be solved by an SVM classifier[2] obtaining a function of the form

$$F(x, y) = \sum_{i=1}^{n} \alpha_i z_i \mathbb{K}(x_i^{(1)}, x_i^{(2)}, x, y) \tag{2}$$

where the pairs $(x_i^{(1)}, x_i^{(2)})$ are the support vectors, and $\mathbb{K}$ is the kernel used. The key idea is the definition of the kernel $\mathbb{K}$ as

$$\mathbb{K}(x_1, x_2, x_3, x_4) = k(x_1, x_3) - k(x_1, x_4) - k(x_2, x_3) + k(x_2, x_4) \tag{3}$$

where k is a kernel function defined as the inner product of the representation of two objects in the features space. In the experiments reported in the next section, we will employ a linear ($k(x, y) = \langle x, y \rangle$) and a polynomial kernel ($k(x, y) = (\langle x, y \rangle + 1)^2$).

3 EXPERIMENTAL RESULTS

To illustrate the benefits of our approach, we have conducted some experiments with a couple of sensory data bases. In both cases we performed a comparison between the scores achieved by preference approaches and those obtained by regression methods. To estimate the errors, we used 10-fold cross validation repeated 5 times.

As was explained above, the core point is the concept of testing session. Thus, for each session, to summarize the opinions of consumers, we computed the mean of the ratings obtained by each food

[1] Artificial Intelligence Center, Gijón (Asturias), Spain. E-mail: {jdiez,gbayon,quevedo,juanjo,oluaces,jalonso,antonio}@aic.uniovi.es

[2] The implementation used in this paper is Joachims' SVM$^{\text{light}}$ [6].

Table 1. Beef meat (above) and cider (below) error results. In regression we report the relative mean absolute deviation; in preferences, the percentage of preference judgments pairs misclassified is shown.

	Regression		Preferences			
	Linear	Cubist	SVM linear	**SVM Poly**	Linear	Cubist
tenderness	96.3%	97.8%	29.6%	19.4%	41.5%	43.1%
flavor	99.3%	103.4%	32.7%	23.8%	43.8%	46.5%
acceptance	94.0%	97.2%	31.9%	22.1%	38.4%	40.2%
Average	96.51%	99.49%	31.39%	**21.79%**	41.24%	43.27%
acidity	103.0%	109.4%	29.9%	18.0%	40.0%	42.4%
bitterness	105.8%	111.9%	30.5%	23.1%	56.0%	47.4%
flavor-1	105.3%	111.7%	27.2%	17.1%	42.4%	44.3%
flavor-2	107.2%	116.0%	28.6%	17.9%	45.6%	45.0%
flavor-3	110.3%	107.7%	33.6%	17.7%	43.8%	41.8%
bouquet	104.0%	110.2%	26.4%	21.0%	43.5%	42.7%
color	98.4%	109.9%	26.1%	17.8%	41.3%	43.4%
visual-1	103.2%	113.0%	25.9%	13.4%	41.7%	43.1%
visual-2	102.3%	112.0%	34.0%	20.0%	43.8%	45.7%
visual-3	107.2%	120.5%	25.3%	20.6%	45.6%	49.3%
visual-4	98.7%	97.2%	23.0%	14.0%	36.5%	38.2%
Average	104.12%	110.87%	28.24%	**18.23%**	43.65%	43.92%

product, which is endowed to the objects' descriptions to conform the regression training sets. Notice that in this context all consumers have tested all products at the same time. Such training set can be used to induce a function that predicts the exact ratings of consumers. We made this experiment with a simple linear regression and with a well reputed regression algorithm: Cubist (RuleQuest Research).

On the other hand, we can obtain some preference judgments comparing the rating of each product with the rest, one by one, and constructing the corresponding ordered pair. To learn from these preference judgment data sets we used SVM with linear and polynomial kernels. In this case, the errors have a straightforward meaning as misclassifications; but in order to allow a fair comparison between regression and preference learning approaches, we also tested regression models on preference judgments test sets, calculating their misclassifications.

The first data base comes from a study carried out to determine the attributes that entail consumer acceptance of beef meat [3] while the second data base deals with sensory data about traditional Asturian cider [10]. Experimental results are shown in Table 1.

First, let us observe that regression methods are unable to learn any useful knowledge: their relative mean absolute deviation ($rmad$) is near 100% in all cases, what means that the mean absolute deviation is more or less the same as that of the mean predictor. Even when we use what was learned with Cubist or a simple linear regression in order to discriminate what was preferred, then the scores are very poor; a cross validation shows that, in average, in this way the errors are over 40%.

On the other hand, when we use the preference learning approach, the usefulness of the models so obtained is considerably higher. Separating methods based on SVM as described in Section 2 can reduce these errors to reach around 30% when we use a linear kernel, but we obtain errors near 20% if the kernel is a polynomial of degree 2. The rationale behind the improvement achieved by nonlinear kernels can be explained taking into account that the positive appreciation of food products usually requires a precise equilibrium of its components, and the increase or decrease of any value from that point is frequently rejected.

4 CONCLUSIONS

The analysis of sensory data is a very interesting issue for food industries, since it provides the knowledge that allows leading production systems in order to satisfy the consumers' sensory requirements. Regression algorithms can not be successfully applied because these methods do not take into account that consumers do not rate all available products; they only assess groups or batches of products presented in a small number of sessions; and consumers give numerical assessments only as a way to express a relative preference, not to be considered as a absolute rating.

Our proposal is to learn functional models able to explain consumer preferences, instead of the exact ratings. In a very practical sense, we can conclude that consumer panels should be asked to concentrate in providing preference judgments pairs instead of lists of ratings. Experimental results show that non-linear functional models achieve the best accuracies in the two real-world sensory data sets analyzed.

ACKNOWLEDGEMENTS

The research reported in this paper has been supported in part under Spanish Ministerio de Ciencia y Tecnología (MCyT) and Feder grant TIC2001-3579. The authors would additionally like to thank: Carlos Sañudo (Animal Production, University of Zaragoza), for providing us with the beef meat data sets [3]; Anna Picinelli (Cider group at SERIDA, Asturias) for allowing us to use the cider data sets [10].

References

[1] J. Díez, J.J. del Coz, O. Luaces, F. Goyache, J. Alonso, A.M. Peña, and A. Bahamonde, 'Learning to assess from pair-wise comparisons', in *Procs. of IBERAMIA*, volume 2527 of *Lecture Notes in Artificial Intelligence*, pp. 481–490, Sevilla, Spain, (2002). Springer-Verlag.

[2] S. Dumais, K. Bharat, T. Joachims, and A. Weigend, eds. *Workshop on implicit measures of user interests and preferences. In ACM SIGIR Conference*, Toronto, Canada, 2003.

[3] M. Gil, X. Serra, M. Gispert, M.A. Oliver, C. Sañudo, B. Panea, J.L. Olleta, M.M. Campo, M. Oliván, K. Osoro, M.D. Garcia-Cachan, R. Cruz-Sagredo, M. Izquierdo, M. Espejo, M. Martín, and J. Piedrafita, 'The effect of breed-production systems on the myosin heavy chain 1, the biochemical characteristics and the colour variables of longissimus thoracis from seven spanish beef cattle breeds', *Meat Science*, **58**, 181–188, (2001).

[4] F. Goyache, A. Bahamonde, J. Alonso, S. López, del Coz J.J., J.R. Quevedo, J. Ranilla, O. Luaces, I. Alvarez, L.J. Royo, and J. Díez, 'The usefulness of artificial intelligence techniques to assess subjective quality of products in the food industry', *Trends in Food Science and Technology*, **12**(10), 370–381, (2001).

[5] R. Herbrich, T. Graepel, and K. Obermayer, *Large Margin Rank Boundaries for Ordinal Regression.*, in *Advances in Large Margin Classifiers*, eds., A.J. Smola, P.L. Bartlett, B. Scholkopf, and D. Schuurmans, 115–132, MIT Press, Cambridge, MA, 2000.

[6] T. Joachims, 'Making large-scale support vector machine learning practical', in *Advances in Kernel Methods: Support Vector Machines*, ed., Smola, Schölkopf, Burges, MIT Press, Cambridge, MA, (1998).

[7] T. Joachims, 'Optimizing search engines using clickthrough data', in *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, (2002).

[8] J.M. Murray, C.M. Delahunty, and I.A. Baxter, 'Descriptive sensory analysis: past, present and future', *Food Research International*, **36**, 461–471, (2001).

[9] T. Næs and E. Risvik, *Multivariate analysis of data in sensory science*, Elsevier, 1996.

[10] A. Picinelli, B. Suárez, J. Moreno, R. Rodríguez, L. Caso-García, and J.J. Mangas, 'Chemical characterization of Asturian cider', *Journal of Agricultural and Food Chemistry*, **48**, 3997–4002, (2000).